

Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla

Jay McCarren and Edward F. DeLong*

Department of Civil and Environmental Engineering and
Division of Biological Engineering, Massachusetts
Institute of Technology, Cambridge, MA 02139, USA.

Summary

Since the recent discovery of retinylidene proteins in marine bacteria (proteorhodopsins), the estimated abundance and diversity of this gene family has expanded rapidly. To explore proteorhodopsin photosystem evolutionary and distributional trends, we identified and compared 16 different proteorhodopsin-containing genome fragments recovered from naturally occurring bacterioplankton populations. In addition to finding several deep-branching proteorhodopsin sequences, proteorhodopsins were found in novel taxonomic contexts, including a betaproteobacterium and a planctomycete. Approximately one-third of the proteorhodopsin-containing genome fragments analysed, as well as a number of recently reported marine bacterial whole genome sequences, contained a linked set of genes required for biosynthesis of the rhodopsin chromophore, retinal. Phylogenetic analyses of the retinal biosynthetic genes suggested their co-evolution and probable coordinated lateral gene transfer into disparate lineages, including *Euryarchaeota*, *Planctomycetales*, and three different proteobacterial lineages. The lateral transfer and retention of genes required to assemble a functional proteorhodopsin photosystem appears to be a coordinated and relatively frequent evolutionary event. Strong selection pressure apparently acts to preserve these light-dependent photosystems in diverse marine microbial lineages.

Introduction

Retinylidene proteins (commonly called rhodopsins) comprise a large and functionally diverse family of proteins.

Different varieties of these photoactive integral membrane proteins are capable of performing two known functions; ion-translocating rhodopsins can act as light-driven ion pumps, and sensory rhodopsins, in the course of light-activated conformational changes, interact with transducers to direct phototaxis and motility (Spudich *et al.*, 2000). Rhodopsins can be broadly categorized as either type I rhodopsins, which are all of microbial origin and found in all three domains of life, or type II rhodopsins, which are all animal photoreceptors (Spudich *et al.*, 2000). Type I rhodopsins include several classes originally identified in extremely halophilic archaea: the two ion-pumping rhodopsins, halorhodopsin (HR) and bacteriorhodopsin (BR), which pump chloride ions and protons, respectively, and two sensory rhodopsins (SRI and SRII). Proteorhodopsins (PR) are another diverse clade of type I rhodopsins first identified in an uncultivated gammaproteobacterium (Beja *et al.*, 2000a), but now known to be prevalent in other bacteria (de la Torre *et al.*, 2003; Sabehi *et al.*, 2005). Recent additions to the large type I rhodopsin family include a monophyletic clade of fungal rhodopsins (Ruiz-Gonzalez and Marin, 2004), another well-supported clade of sequences of unknown function from diverse microbes, and several sequences representing independent branches in the type I rhodopsin phylogenetic trees (Fig. 1).

Originally, microbial rhodopsins were typically associated with *Archaea*, until a PR was identified in an uncultured gammaproteobacterium of the SAR86 clade (Beja *et al.*, 2000a). Genome fragments from other members of the SAR86 group from diverse geographic locales, and other taxa including *Alphaproteobacteria* (de la Torre *et al.*, 2003), were also shown to encode the photoprotein (Sabehi *et al.*, 2003). Now, PR-like sequences number in the hundreds (Venter *et al.*, 2004), and PRs are estimated to be present in 13% of all marine bacteria in the photic zone (Sabehi *et al.*, 2005). Surprisingly, recent metagenomic analyses revealed the presence of marine bacterial-like PRs within the genomes of marine planktonic *Euryarchaea* (Frigaard *et al.*, 2006). Conversely, the complete genome sequence of the extremely halophilic bacterium *Salinibacter ruber* contained several archaeal-like rhodopsins (HR, SRI and SRII), similar to those from its archaeal neighbours that dwell in the same hypersaline

Received 21 September, 2006; accepted 2 November, 2006. *For correspondence. E-mail delong@mit.edu; Tel. (+1) 617 253 0252; Fax (+1) 617 253 2679.

Type I Rhodopsins

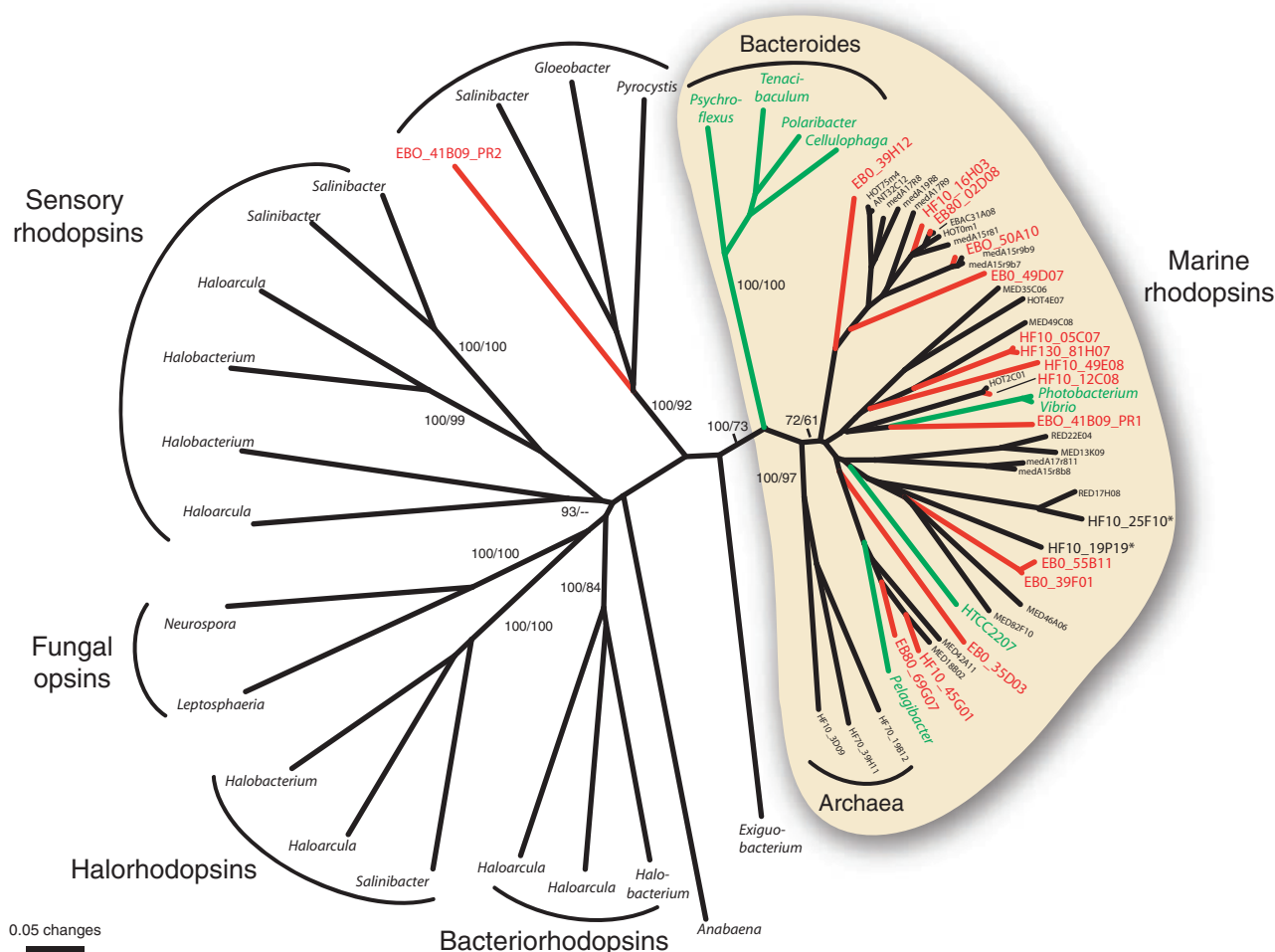


Fig. 1. Neighbour-joining tree (of inferred amino acid sequences) containing representatives from the major groups of type I rhodopsins. New sequences reported in this study are indicated in red [the asterisk (*) indicates sequences reported by A. Martinez and E.F. DeLong, in preparation] and sequences from recently reported marine bacterioplankton genome sequencing (<http://www.moore.org/microgenome/>) are indicated in green. Bootstrap support values from 1000 pseudoreplicates given for major nodes (neighbour-joining/maximum parsimony; --, unresolved node).

habitat (Mongodin *et al.*, 2005). The distribution of these ubiquitous photoproteins has been elucidated through analysis of both cultivated isolates (Kaneko *et al.*, 2001; Sineshchekov *et al.*, 2002; Galagan *et al.*, 2003; Jung *et al.*, 2003; Nakamura *et al.*, 2003; Okamoto and Hastings, 2003; Giovannoni *et al.*, 2005a) (<http://www.moore.org/microgenome/>) and through culture-independent approaches (Beja *et al.*, 2000a; 2001; Sabehi *et al.*, 2003; 2004; 2005; de la Torre *et al.*, 2003; Frigaard *et al.*, 2006). These reports provide examples of the broad distribution of different type I rhodopsins found among all three domains of life.

Two components are required for a functional, photo-active rhodopsin: the opsin apoprotein itself and its chromophore retinal. Perhaps not surprisingly then, recent

sequencing of large insert bacterial fosmids that contain proteorhodopsin revealed the presence of retinal biosynthesis genes alongside the PR gene in several clones (Sabehi *et al.*, 2005). Retinal is an apocarotenoid formed by the oxidative cleavage of β -carotene (Giuliano *et al.*, 2003). With the exception of most oxyphotoautotrophic bacteria and plants, β -carotene synthesis from isoprenoid precursors is accomplished with the four *crt* genes *crtE*, *crtI*, *crtB* and *crtY* (Armstrong, 1997). Oxidative cleavage of β -carotene yields the rhodopsin chromophore retinal (Giuliano *et al.*, 2003). The gene products of *blh* and *brp* have been shown to have this activity in *Halobacterium salinarum* (Peck *et al.* 2001), while expression of a homologous gene from PR-containing clones had the same activity in *Escherichia coli* (Sabehi *et al.*, 2005).

Table 1. Summary of proteorhodopsin-containing environmental clones analysed in this study.

Clone ^a	Putative origin	Evidence ^b	Retinal biosynthesis
EB0_35D03	<i>Proteobacterium</i> ^c	CBHs (53.8%)	<i>crtEIBY, blh</i>
EB0_39F01	<i>Alphaproteobacterium</i>	CBHs (88.6%)	
EB0_39H12	<i>Proteobacterium</i> ^c	CBHs (80.7%)	
EB0_41B09	<i>Betaproteobacterium</i>	Phylogeny of three ribosomal proteins	
EB0_49D07	<i>Proteobacterium</i> ^c	CBHs (60.%)	<i>crtEIBY, blh</i>
EB0_50A10	<i>Gammaproteobacterium</i>	CBHs (60.9%)	
HF10_05C07	<i>Proteobacterium</i>	CBHs (88.1%)	
HF10_12C08	<i>Alphaproteobacterium</i>	CBHs (73.5%)	
HF10_19P19 ^d	<i>Proteobacterium</i> ^e	Phylogeny of four ribosomal proteins	<i>crtEIBY, blh, idi</i>
HF10_25F10 ^d	<i>Proteobacterium</i> ^e	CBHs (91.9%)	
HF10_29C11 ^f	<i>Euryarchaea</i>	CBHs (64.7%)	
		RNA polymerase A'' phylogeny	
HF10_45G01	<i>Proteobacterium</i> ^e	CBHs (92.7%)	<i>crtEIBY, blh, idi</i>
HF10_49E08	Planctomycete	CBHs (38.5%)	
		CBHs (47.4%)	
		XylA phylogeny	
EB80_02D08	SAR-86	CBHs (65.8%)	<i>crtEIBY, blh</i>
EB80_69G07	<i>Alphaproteobacterium</i>	16S rRNA	
HF130_81H07	<i>Gammaproteobacterium</i>	CBHs (89.7%)	
		CBHs (50.0%)	

a. Characters preceding '_' indicate location (HF, Hawaiian fosmid; EB, Monterey Bay BAC) and depth (0 m, 10 m, 80 m and 130 m) of sample collection.

b. Percentage of ORFs matching listed origin based on CBHs over threshold e -value of e^{-10} . When CBH was an uncultivated environmental clone, the next best hit with an identified taxonomic affiliation was used.

c. CBHs spread among proteobacterial taxa precluding assignment to a specific class.

d. Sequences reported by A. Martinez and E.F. DeLong (in preparation)

e. CBHs match uncultivated PR-containing proteobacterial clones.

f. Proteorhodopsin was not linked to retinal biosynthesis genes in *Euryarchaea*.

CBH, closest BLASTP homologue; ORFs, open reading frames.

These retinal biosynthetic genes have been found immediately downstream of PR in a number of large insert BAC and fosmid clones. In a few instances *idi* genes that catalyse isomerization of the isoprenoid precursors (Kaneda *et al.*, 2001; Kuzuyama and Seto, 2003) and enhance the yield of β -carotene (Kajiwara *et al.*, 1997; Sedkova *et al.*, 2005) have also been found adjacent to these gene clusters.

Through the analysis of large fragments of chromosomal DNA the first PR sequence was discovered (Beja *et al.*, 2000a) and continued prospecting of large insert environmental DNA libraries has extended the known distribution of PR-containing organisms (Sabehi *et al.*, 2005; Frigaard *et al.*, 2006). The recent full genome sequencing of a large number of marine bacterial isolates has also added to the variety of microorganisms known to possess PR sequences (<http://www.moore.org/microgenome/>). In light of the apparently common lateral transfer of PR genes, determining the origin of a proteorhodopsin based exclusively on its sequence without additional genetic contextual data is problematic. We are then left with sequencing whole genomes, or large chromosomal fragments, to ascertain their origins. We present here an extended survey and analysis of 16 large insert clones, recovered from naturally occurring bacterioplankton populations, that possess PR-like sequences.

Results

Proteorhodopsin sequence diversity

Thirteen large insert DNA clones were identified by colony hybridization of macroarrays using amplified PR genes as probes, and subsequently sequenced. Additionally, two fully sequenced fosmid clones that contain proteorhodopsin photosystem genes were identified by expression analysis (A. Martinez and E.F. DeLong, unpublished) and are also included in these analyses. All PR-containing clones originated from either Monterey Bay (EB prefix) or North Pacific Subtropical Gyre (HF prefix) euphotic zone waters, between 0 and 130 m depth (Table 1). Phylogenetic analysis of the PR sequences indicated that all fell into clades of previously reported polymerase chain reaction (PCR)-amplified or shotgun clone PRs (Venter *et al.*, 2004; Sabehi *et al.*, 2005) with one notable exception (Fig. 1). A few of the large genome fragments reported here were nearly identical in gene content, order and amino acid sequence to previously reported PR-containing clones. Many of the BAC or fosmid-encoded PRs, however, represented divergent subclades within the marine microbial PR tree.

Surprisingly, BAC clone EB0_41B09 contained two tandemly arranged PR-like open reading frames (ORFs) (Fig. 2). While EB0_41B09_PR1 grouped unambiguously within the large marine bacterial PR clade,

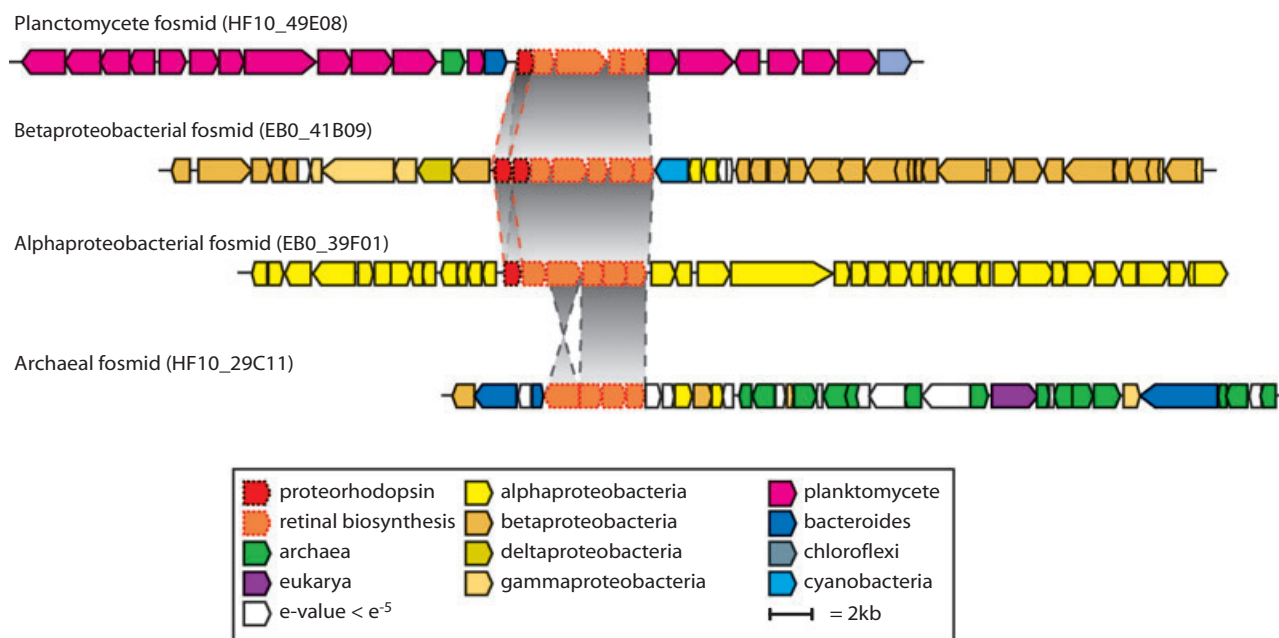


Fig. 2. Gene organization of four environmental fosmid clones containing retinal biosynthesis genes and, in three instances, proteorhodopsin genes. Homologous regions are indicated by shaded regions. Genes flanking homologous regions are colour coded according to phylogenetic affiliation of best BLAST hit.

EB0_41B09_PR2 is quite divergent (Fig. 1). EB0_41B09_PR2 belonged to a well-supported clade containing other PR-like sequences from the genomes of *S. ruber* (Mongodin *et al.*, 2005) and *Gloeobacter violaceus* (Nakamura *et al.*, 2003) and a gene identified from a *Pyrocystis lunula* cDNA microarray analysis (Okamoto and Hastings, 2003). The PR-like protein from *S. ruber*, named xanthorhodopsin, has been shown to act as a light-activated proton pump (Balashov *et al.*, 2005). Most unusually, EB0_41B09_PR2 did not contain the nearly universally conserved (Spudich *et al.*, 2000) lysine residue, with which retinal forms a covalent Schiff base to form the functional photoprotein. Consequently, whether EB0_41B09_PR2 can bind retinal remains to be determined.

Phylogenetic distribution of proteorhodopsin genes on large genome fragments

As an initial step, top scoring BLASTP returns for each predicted ORF in a given clone were tabulated. Conserved, phylogenetically informative genes in clones (where they were available) were then aligned to a set of homologues and phylogenetic analyses performed (see below). In each case where phylogenetic marker genes were present and analysed, there was good agreement with the putative taxonomic assignments based on cumulative BLASTP returns (Nesbo *et al.*, 2005). Based on these data, 11 clones are of putatively proteobacterial origin,

while several can be assigned more specifically to either the gamma- or alphaproteobacterial classes.

One BAC clone, EB80_02D08, contained a SSU rRNA gene that was nearly identical to the original proteorhodopsin-containing clone, EBAC_31A08, having only a single-base mismatch placing this clone within of the SAR82-II cluster of *Gammaproteobacteria* (Suzuki *et al.* 2001). In clones without rRNA genes, it was sometimes possible to use ribosomal protein sequences to assign taxonomic origins (de la Torre *et al.*, 2003; Teeling and Gloeckner, 2006). Clones HF10_12C08 and EB0_41B09 each contained multiple ribosomal protein genes, and concatenated ribosomal protein sequence analyses yielded well-supported trees, allowing preliminary assignment of each of these PRs to a specific taxon (Figs S1 and S2 in *Supplementary material*). A phylogenetic tree based on four ribosomal proteins (L1, L10, L11 and L7/L12) present on clone HF10_12C08 confirmed that this clone was derived from an alphaproteobacterium. The closest homologues of these four proteins are from another PR-containing alphaproteobacterial clone, HOT2C01 (de la Torre *et al.* 2003). HF10_12C08 and HOT2C01 share an extensive region of homology (approximately 88% DNA identity over 28 kb), which also contained other phylogenetically informative genes including translation elongation factor Tu and RNA polymerase beta subunit (*rpoB*) in addition to the ribosomal proteins listed. Clone EB0_41B09 also contained genes for a number of ribosomal proteins (S4, S6, S18 and L9).

Concatenated ribosomal protein alignments of the three ribosomal proteins S6, S18 and L9 produced a well-supported phylogenetic tree that indicates a betaproteobacterial origin for clone EB0_41B09, a novel taxonomic origin for a proteorhodopsin. The most similar homologues for nearly two-thirds of the ORFs encoded on EB0_41B09 are of betaproteobacterial origin (Fig. 2). Virtually all ORFs that were not most similar to betaproteobacterial homologues clustered together on the chromosome immediately flanking the PR gene. This cluster of genes is bounded on one end by a tRNA gene, which are frequent insertion sites for lateral gene transfer (LGT), and may represent an insertion into a betaproteobacterial background.

Seventeen of the 26 predicted ORFs from clone HF10_49E08 shared high similarity to homologues from planctomycete-derived protein sequences (Fig. 2). Moreover, six ORFs encoded putative sulfatases, known to be highly over-represented in the genome of the planctomycete *Pirellula* sp. strain 1 (Glockner *et al.*, 2003). HF10_49E08 also possessed a xylose isomerase (*xyIA*) gene, which appeared to be well conserved across a wide range of bacteria. Phylogenetic analyses of XylA protein sequences largely recovered the monophyly of included taxa, and unambiguously grouped the HF10_49E08 XylA with the other two planctomycetes (Fig. S3 in *Supplementary material*). Taken together, these data indicate that members of the *Planctomycetales*, like other planktonic marine bacteria, contain proteorhodopsin photosystems.

We searched for the retinal biosynthetic genes in planktonic marine *Euryarchaea*, as they did not appear to be linked to archaeal PRs in a previous study (Frigaard *et al.*, 2006). Clone HF10_29C11 (which does not contain a PR gene, but encodes several retinal biosynthesis genes, see below) contained a complete RNA polymerase A' subunit gene (*rpoA2*), which are of some utility in archaeal phylogenetic analysis (Brochier *et al.*, 2004). The *rpoA2* from HF10_29C11 is clearly of euryarchaeal origin based on our phylogenetic analysis, although its exact position within the group is poorly resolved (Fig. S4 in *Supplementary material*). HF10_29C11 also contained a partial *rpoA1* gene and two genes encoding ribosomal proteins, all of which support a euryarchaeal origin of this clone. Further supporting its archaeal origins, HF10_29C11 encodes a conserved hypothetical ORF that is present in nearly two-thirds of all *Archaea* sequenced to date and has yet to be found outside of the archaeal domain.

Genome sequencing of marine bacterioplankton isolates has revealed the presence of PR-like sequences in a number of isolates currently in culture. Two *Pelagibacter ubique* strains of the abundant SAR11 clade of *Alphaproteobacteria* possess a PR (Giovannoni *et al.*, 2005b) (<http://www.moore.org/microgenome/>). Three cultivated

Gammaproteobacteria (*Photobacterium* sp. strain SKA34, *Vibrio angustum* strain S14, and marine gammaproteobacterial strain HTCC2207), as well as several *Bacteroidetes* isolates, can now be added to the list of cultured PR-containing marine microbes, based on data from the recent Moore Foundation Marine Microbial Genome Sequencing Project (<http://www.moore.org/microgenome/>). The proteorhodopsins from the flavobacteria-like *Bacteroidetes*, including *Cellulophaga*, *Polaribacter*, *Psychroflexus* and *Tenacibaculum* species, cluster together in a separate deep-branching clade compared with previously reported proteorhodopsins (Fig. 1). Unlike most proteorhodopsins identified to date, the *Bacteroidetes* PRs encode a methionine residue at position 105, a site shown to be largely responsible for tuning the wavelength of light absorbed by PR (Man *et al.*, 2003), although other residues within the PR molecule are putatively involved in spectral fine tuning (Bielawski *et al.*, 2004).

Proteorhodopsin-linked photopigment biosynthetic genes

Highly similar proteorhodopsins in different large insert clones often were flanked by similar sets of genes. The gene content and order around closely related proteorhodopsin sequences in large insert clones frequently shared extensive regions of DNA homology, suggesting that these clones originated from closely related organisms. Comparisons of PR flanking regions from more distantly related proteorhodopsins generally (but not always, see below) revealed few commonly linked genes. Aside from the PR itself, no genes were found common to all or even most of these clones. However, one trend was observed in searching for additional commonalities between these clones. Nearly one-third of the large insert PR-containing DNA clones from marine surface water microbes (Table 1) encoded all the genes required for retinal biosynthesis from isoprenoid precursors. The five genes involved in retinal biosynthesis were encoded on the same DNA strand as the PR gene, and were located immediately downstream, similar to that shown for a few other PR-containing environmental clones (Sabehi *et al.*, 2005). Furthermore, the clones containing linked retinal biosynthesis genes spanned a broad phylogenetic range, including *Alpha*-, *Beta*- and *Gammaproteobacteria*, and *Planctomycetales*. Marine planktonic euryarchaeal-associated retinal biosynthetic genes were also identified, but these were not closely linked to PR.

The carotenoid biosynthesis genes *crtE*, *crtB*, *crtI* and *crtY* encode the enzymes necessary for the synthesis of β -carotene (Sandmann, 1994), and *blh* cleaves this carotenoid to yield retinal (Sabehi *et al.*, 2005). Gener-

ally, these genes are arrayed downstream of the PR gene in the putative operonal arrangement *crtEIBY*, *blh* (Fig. 2, Table 1). Clones HF10_19P19 and HF10_25F10 possess an additional downstream ORF similar to IPP isomerase (*idi*) to give the arrangement *crtEIBY*, *blh*, *idi*. IPP isomerases catalyse the interconversion of the isoprenoid precursors required for β -carotene, and have been shown to enhance carotenoid production (Kajiwara *et al.*, 1997; Sedkova *et al.*, 2005). Notably *crtI* and *crtB* in clone HF10_49E08 appear to have undergone a gene fusion and are present as a single ORF. Finally the archaeal clone HF10_29C11 lacks a *crtE* gene, and the *crtI* gene is encoded on the opposite strand compared with its other retinal biosynthesis gene homologues (Fig. 2).

Phylogenetic trees were constructed for each gene involved in retinal biosynthesis (Figs 3 and 4, and Fig. S5 in *Supplementary material*). For each tree, several strongly supported clades were present, which generally corresponded to the clone's presumptive overall taxonomic origin. There were interesting exceptions, however, where chromophore biosynthetic gene relationships disagreed with taxonomic origins. The clearest illustration of this could be seen in carotenoid biosynthetic genes derived from the planctomycete and euryarchaeal large insert DNA clones. The phylogenetic trees for CrtE, CrtB and CrtI all show a strongly supported planctomycete clade (*Blastopirellula* and *Rhodopirellula* homologues), while the planctomycete clone HF10_49E08 carotenoid biosynthetic genes group with the clade of marine PR-containing microbes, representing mostly *Proteobacteria*. Likewise, CrtB and CrtY trees strongly supported specific archaeal clades distinct from bacterial homologues. The exception was the archaeal clone HF10_29C11, whose CrtB and CrtY were affiliated with the marine PR-containing clade. This is similar to the relationship observed for the euryarchaeal-associated PR that is phylogenetically closely related to the PRs of marine planktonic bacteria (Frigaard *et al.*, 2006). In contrast, CrtE, CrtB and CrtI from PR-containing *Bacteroidetes* genomes formed an independent clade. In only one case (CrtY) did *Bacteroidetes* sequences cluster with the PR-linked proteobacterial clade (Fig. 4B). Phylogenies for CrtB, CrtY and Blh also reveal that the halophilic *Bacteroidetes* bacterium *S. ruber* possesses several retinal biosynthesis enzymes more similar to halophilic archaea than to other *Bacteroidetes*.

In addition to two tandem proteorhodopsin genes and the adjacent retinal biosynthetic pathway genes, the betaproteobacterial clone EBO_41B09 also contained additional ORFs with homology to other light-activated genes. Approximately 300 bases upstream from the start site of EBO_41B09_PR1, and encoded on the opposite DNA strand an ORF bearing high similarity to DNA pho-

tolyases was present. At the opposite end of the PR photosystem gene cluster, and encoded on the opposing strand, is a gene bearing high similarity to bacterial cryptochromes (Hitomi *et al.*, 2000). DNA photolyases and cryptochromes share significant sequence homology (Cashmore *et al.*, 1999), which suggests potentially similar mechanisms of light energy activation (Essen, 2006). The exact function of bacterial cryptochromes, however, remains to be determined.

Discussion

Proteorhodopsins, a class of retinylidene proteins only recently discovered, are rapidly becoming recognized as widespread, diverse and abundant photoproteins found in diverse ecological and genomic contexts. A variety of approaches have helped reveal their distribution and variability including environmental PCR surveys (Man *et al.*, 2003; Sabehi *et al.*, 2003), directed sequencing of BAC of fosmid clones (Beja *et al.*, 2000a; Sabehi *et al.*, 2003; 2004; 2005; de la Torre *et al.*, 2003; Frigaard *et al.*, 2006; A. Martinez and E.F. DeLong, unpublished), whole genome sequencing (Giovannoni *et al.*, 2005b; Mongodin *et al.*, 2005; <http://www.moore.org/microgenome/>), and random shotgun sequencing of environmental DNA (Venter *et al.*, 2004; DeLong *et al.*, 2006). This study provides comparative perspective on the phylogenetic distribution of PR genes, their genetic linkage and evolutionary relatedness, and the likely co-transfer, retention and co-evolution of the proteorhodopsin photosystem gene cluster.

For several years proteorhodopsins were only reported in uncultivated members of the *Gamma*- and *Alphaproteobacteria* (Man *et al.*, 2003; Sabehi *et al.*, 2003; 2004; de la Torre *et al.*, 2003). Proteorhodopsins are now known to be much more widely distributed. For example, in the North Pacific Subtropical Gyre, proteorhodopsins were apparently present in most of the euryarchaeal population in the euphotic zone (Frigaard *et al.*, 2006). Additionally, proteorhodopsins are present within the genome of both sequenced strains of the abundant SAR11 clade of *Alphaproteobacteria* (Giovannoni *et al.*, 2005a,b) (<http://www.moore.org/microgenome/>). While the majority of the clones we report here do appear to originate from either *Gamma*- or *Alphaproteobacteria*, two clones are derived from marine bacterioplankton taxa not previously reported to harbour proteorhodopsins. These include putative members of the *Betaproteobacteria* and the *Planctomycetales*, both groups known to be well represented in marine bacterioplankton, especially in the coastal regions from which these genome fragments originate (Suzuki *et al.*, 2004). Shotgun sequencing had earlier hinted (Venter *et al.*, 2004), and full genome sequencing of isolates now has confirmed (<http://www.moore.org/microgenome/>), that

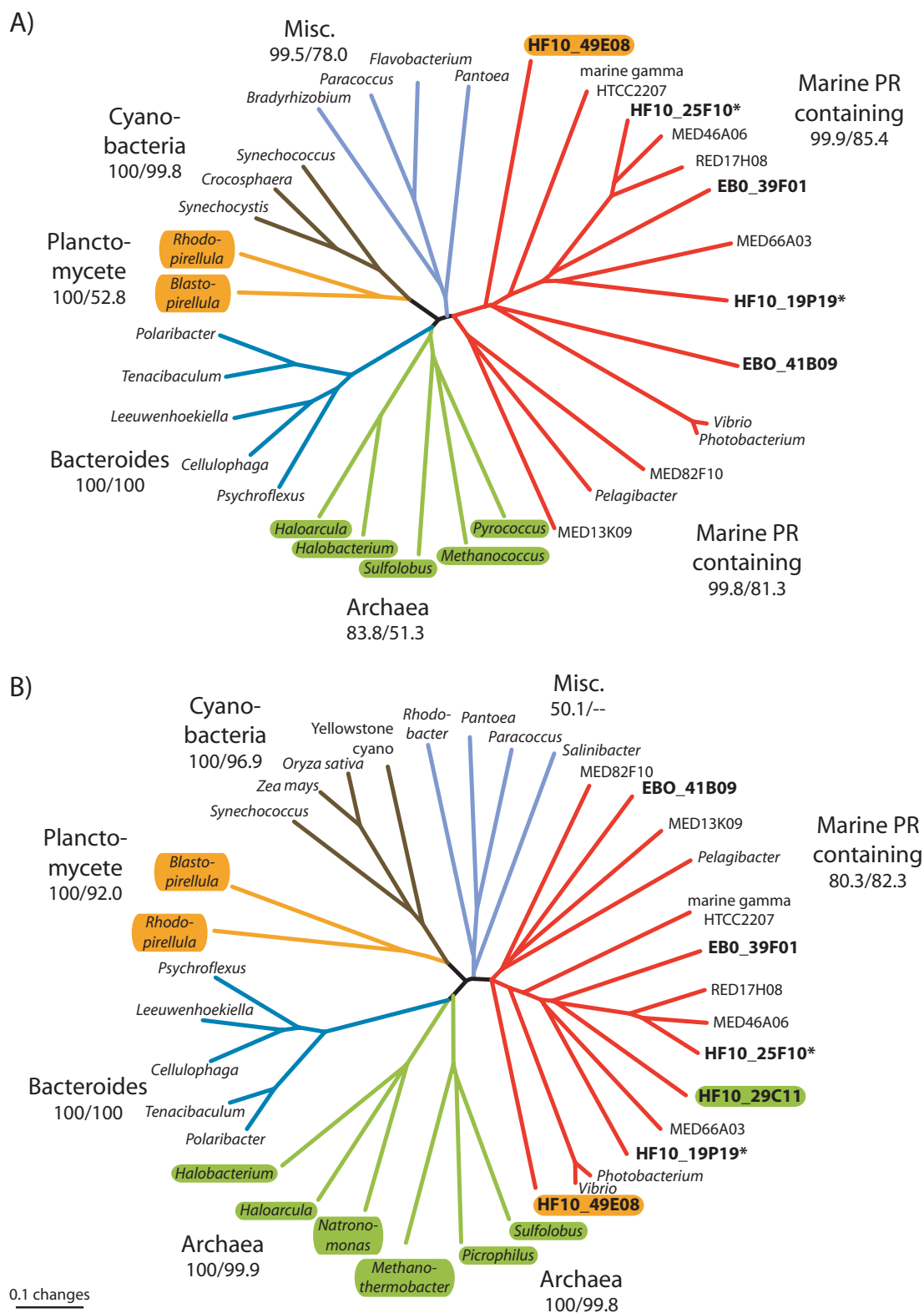


Fig. 3. Phylogenetic trees of carotenoid biosynthesis enzymes (A) CrtE and (B) CrtB based on comparisons of 336 and 317 parsimony informative characters respectively. Branches of major clades are differentiated by colour with bootstrap values of 1000 pseudoreplicates for these clades given (neighbour-joining/maximum parsimony; --, unresolved node). Newly deposited sequences marked in bold [the asterisk (*) indicates sequences reported by A. Martinez and E.F. DeLong, in preparation]. Coloured boxes around individual sequence names highlight those that group outside of their taxonomic clade.

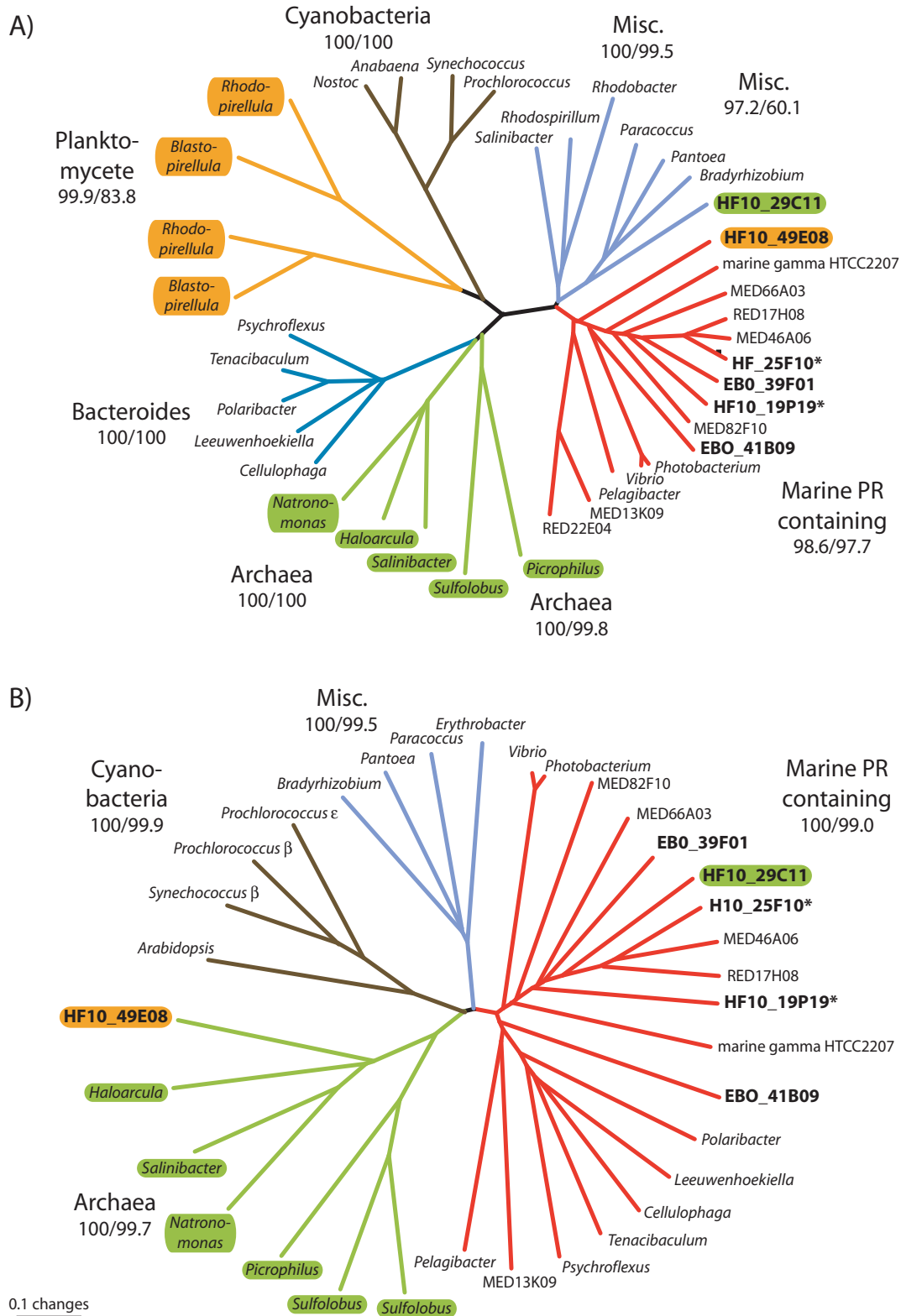


Fig. 4. Phylogenetic trees of carotenoid biosynthesis enzymes (A) CrtI and (B) CrtY based on comparisons of 519 and 409 parsimony informative characters respectively. Branches of major clades are differentiated by colour with bootstrap values of 1000 pseudoreplicates for these clades given (neighbour-joining/maximum parsimony). Newly deposited sequences marked in bold [the asterisk (*) indicates sequences reported by A. Martinez and E.F. DeLong, in preparation]. Coloured boxes around individual sequence names highlight those that group outside of their taxonomic clade.

many marine *Bacteroidetes* (flavobacteria-related) also encode PR-like genes. Interestingly, even commonly cultivated taxa, such as marine *Vibrio* or *Photobacterium* species, also encompass strains that have acquired this photoprotein. Proteorhodopsins are widely distributed across the universal tree of life [e.g. proteorhodopsin-like sequences are present in the eukaryote *P. lunula* (Okamoto and Hastings, 2003), in marine group II euryarchaeotes (Frigaard *et al.*, 2006) and widely distributed among the eubacteria]. It is likely that the list of taxa containing PRs will continue to expand with continued exploration of genomes and cultivars.

Multiple lines of evidence suggest that proteorhodopsins are relatively frequently transferred between different lineages. For example, proteorhodopsins are found widely dispersed across diverse organisms, and proteorhodopsin phylogenies do not generally correspond to phylogenies for highly conserved genes. While similar PR sequences are found in quite different organisms, the converse is also true with closely related organisms possessing disparate PR sequences (Sabehi *et al.*, 2004). Additionally, and similar to the case for the archaeal bacteriorhodopsin family (Baliga *et al.*, 2004), there are now several instances where a member of the *Bacteria* has been shown to possess multiple, divergent rhodopsin sequences. Similarly, clone HF10_41B09 has two very different PR sequences arrayed in tandem. It appears that rhodopsin gene duplication and diversification, as well as lateral transfer, is a common phenomenon. The different varieties of PR generated by these processes appear to be maintained by considerable positive selection for their retention and utilization in the photic zone.

Without the ability to produce the chromophore retinal, lateral acquisition of a PR gene is functionally irrelevant. To gain selective advantage from PR, a recipient cell needs to already possess retinal biosynthesis capabilities, or to acquire this capacity along with the PR gene. The acquisition of only six genes is required to gain the ability to harvest biochemical energy from light. This modest requirement, for such a bioenergetic advantage, seems to have favoured the widespread acquisition and utilization of the PR photosystem. A previous report (Sabehi *et al.*, 2005) showed that several environmental clones had the genes necessary to biosynthesize retinal, immediately downstream of proteorhodopsin. Many of the environmental clones we report here, although from disparate sources, have a similar chromosomal arrangement, with the PR gene immediately upstream of the β -carotene biosynthetic pathway genes and the gene encoding its cleavage to retinal. The close linkage of these genes suggests that they may all have been acquired together (Wolf *et al.*, 2001). Although mosaic operons have been observed (Omelchenko *et al.*, 2003), the conservation of

gene order across distantly related taxa is strongly suggestive of a laterally transferring operon (Koonin *et al.*, 2001). Phylogenetic analysis of the β -carotene biosynthesis *crt* genes from these clones also indicates that these genes were acquired laterally. For example, a number of *crt* genes (from a planctomycete and an archaeon) were not associated with homologues of their parent taxa, but rather clustered with those of other PR-containing bacteria. A parallel example is also evident in the retinal biosynthesis gene phylogenies of *crtI* and *crtY* from the bacterium *S. ruber* that grouped with halophilic archaea and not with the other *Bacteroidetes*, similar to previous observations for *blh* (Mongodin *et al.*, 2005).

Nearly a third of the PR-containing environmental genome fragments from diverse sources reported here contained PR-linked retinal biosynthetic genes. Although not all PRs were linked to retinal biosynthesis genes, it is probable these genes reside elsewhere within the genome. Among the sequenced marine microbial genomes that contain a single copy of proteorhodopsin, some have the *crt* genes immediately downstream of PR, while others do not. Phylogenetic trees for the β -carotene biosynthetic genes also provide clues about differential gene linkage and transfer that appear, in part, dependent on the recipient's genetic context. Marine *Flavobacteria*, for example, are capable of producing β -carotene (Bernardet *et al.*, 2002) and their β -carotene biosynthesis enzymes are generally quite distinct from the other marine proteorhodopsin-containing organisms. However, the *crtY* gene represents an exception, and highlights the fact that the individual genes within this pathway do not necessarily share a common evolutionary history (Sandmann, 2002). In the genetic context of a β -carotene-synthesizing microorganism, only two genes (*pop* and *blh*) are required to assemble a functional PR photosystem. Presumably the marine *Flavobacteria* would only need to acquire a single gene to synthesize retinal. Consistent with this, in each of the PR-containing marine *Flavobacteria* genomes, while the *crt* genes are not linked to the PR, a *blh* homologue is present immediately flanking the PR gene. Further, a *blh* homologue is absent from the genome of the marine flavobacterium *Leeuwenhoekiella blandensis* str. MED217, which possesses β -carotene biosynthesis genes similar to the other marine *Flavobacteria* (Figs 3 and 4) yet lacks a proteorhodopsin.

Proteorhodopsins are broadly distributed across diverse marine microbial taxa. While this distribution is partly a consequence of LGT, it also probably reflects a substantial fitness advantage conferred by the photosystem. As the enzymatic machinery for synthesizing a complete proteorhodopsin-based photosystem is so simple, requiring only six linked genes, there is a reasonably high probability of transferring a fully functional gene complement. In a single LGT event, a microorganism

might easily acquire the ability to derive energy from sunlight. As a result, the relatedness of coupled rhodopsin and retinal biosynthesis genes seems more closely associated with function and environmental context, than to the particular taxonomic relationships of the organisms from which they are derived. A similar trend is observed in hypersaline environments where the bacterium *S. ruber* and halophilic *Archaea* appear to have exchanged homologous, 'archaeal' elements of the same photosystem. Gene transfer among distantly related taxa inhabiting the same environment has been proposed as one of the salient trends of LGT (Beiko *et al.*, 2005). This appears to be the case for PR-based photosystems, and is likely driven by a strong selection for energy acquisition using light, in nutrient limited environments. At least 10% of all *Bacteria* and *Archaea* in oceanic surface waters appear to contain this photosystem (Sabehi *et al.*, 2005; Frigaard *et al.*, 2006). In the case of marine euryarchaeal genomes, the photoprotein gene is absent in *Euryarchaea* living below the euphotic zone (Frigaard *et al.*, 2006). The proteorhodopsin photosystem appears to represent an important functional component in the marine photic zone 'habitat genome' (defined as the pool of genes that are beneficial for adaptation to a given set of environmental constraints; Mongodin *et al.*, 2005). Moreover, the identification of additional PR-linked genes encoding putative 'photoactive' products (DNA photolyase and cryptochrome from EBO_41B09) suggests potential co-transfer of other light-related genes within the marine euphotic zone. Proteorhodopsins are found broadly distributed among diverse taxa in the microbial plankton, and are dynamic elements in gene ecology. It will be important to define their various functional roles and bioenergetic contributions to better understand the ecological significance of proteorhodopsin photophysiology in the sea, and elsewhere.

Experimental procedures

Fosmid and BAC library construction cloning

The sampling and construction of the fosmid (DeLong *et al.*, 2006) and BAC (Beja *et al.*, 2000b) libraries have been previously described. Large insert DNA clones containing the PR genes were identified by colony hybridization macroarrays, as previously described (de la Torre *et al.*, 2003; DeLong *et al.*, 2006).

Sequencing and annotation

Sequencing was performed by shotgun cloning of and sequencing of large insert clones, as previously described (Hallam *et al.*, 2004; 2006), with the exception of two clones: HF10_49E08 and HF10_29C11. These two fosmid clones were sequenced by first constructing a random

transposon insertion library for each clone using the Tn5 vector EZ-Tn5<KAN> (Epicentre, Madison WI, USA). Tn5 insertion clones were sequenced in both directions using KAN-2 FP-1 and KAN-2 RP-1 primers (Epicentre, Madison, WI, USA) using BigDye v3.1 cycle sequencing chemistry and run on an ABI Prism 3700 DNA analyser (Applied Biosystems, Forest City, CA, USA). Sequences were assembled with Sequencher v4.5 (Gene Codes, Ann Arbor, MI, USA) and annotated with both FGENESB (Softberry, Mount Kisco, NY, USA) and Artemis release 7 (Rutherford *et al.*, 2000).

Library probing

North Pacific Subtropical Gyre fosmid library clones from 0, 70 and 130 m depths were arrayed on Performa II filters (Genetix, Boston, MA, USA) and hybridized with a probe targeting a carotenoid biosynthetic gene. The probe, a 509 bp fragment of a *crtY* gene, was PCR amplified from clone HF10_31E03, gel purified and AlkPhos labelled (Amersham, Piscataway, NJ, USA). Following overnight hybridization at 60°C, membranes were incubated with ECF substrate (Amersham) and visualized on a Fujifilm FLA 5100 scanner.

Sequence analyses

BLAST analyses (Altschul *et al.*, 1997) were used to compare complete clone sequences with results visualized using the custom perl script blastView3 (J. Chapman © 2005) (BLASTN settings $X = 150$, $q = -1$, $F = F$) and the Artemis Comparison Tool (Carver *et al.*, 2005). Ribosomal proteins sequences were extracted from the RibAlignDB database concatenated and their amino acid sequences were aligned using RibAlign software (Teeling and Gloeckner, 2006). All other alignments were generated from inferred amino acid translations using CLUSTALX v1.83.1. Phylogenetic trees were constructed using both neighbour-joining distance methods and maximum parsimony heuristics, performed with PAUP v4.0b10 (Sinauer Assoc., Sunderland, MA, USA). Tree topologies for both maximum parsimony and neighbour-joining trees were tested with bootstrap analyses of 1000 pseudoreplicates using the TBR branch-swapping algorithm.

Accession numbers

Annotated BAC and fosmid sequences have been deposited to the GenBank/EMBL/DDJB databases under Accession No. EF089397–EF089402, EF107099–EF107106, EF100190 and EF100191.

Acknowledgements

We thank Lynne Christianson for assistance with macroarray hybridization experiments. Special thanks to Asuncion Martinez for sharing unpublished sequence data. Thanks also to the JGI staff for assistance in sequencing. This work was supported by a grant from the Gordon and Betty Moore Foundation (E.F.D.), an NSF Microbial Observatory award (MCB-0348001) and a US Department of Energy Microbial Genomes Program award (E.F.D.).

References

- Altschul, S.F., Madden, T.L., Schaeffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Armstrong, G.A. (1997) Genetics of eubacterial carotenoid biosynthesis: a colorful tale. *Annu Rev Microbiol* **51**: 629–659.
- Balashov, S.P., Imasheva, E.S., Boichenko, V.A., Anton, J., Wang, J.M., and Lanyi, J.K. (2005) Xanthorhodopsin: a proton pump with a light-harvesting carotenoid antenna. *Science* **309**: 2061–2064.
- Baliga, N.S., Bonneau, R., Facciotti, M.T., Pan, M., Glusman, G., Deutsch, E.W., *et al.* (2004) Genome sequence of *Haloarcula marismortui*: a halophilic archaeon from the Dead Sea. *Genome Res* **14**: 2221–2234.
- Beiko, R.G., Harlow, T.J., and Ragan, M.A. (2005) Highways of gene sharing in prokaryotes. *Proc Natl Acad Sci USA* **102**: 14332–14337.
- Beja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P., *et al.* (2000a) Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**: 1902–1906.
- Beja, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A., Nguyen, L.P., *et al.* (2000b) Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* **2**: 516–529.
- Beja, O., Spudich, E.N., Spudich, J.L., Leclerc, M., and DeLong, E.F. (2001) Proteorhodopsin phototrophy in the ocean. *Nature* **411**: 786–789.
- Bernardet, J.F., Nakagawa, Y., and Holmes, B. (2002) Proposed minimal standards for describing new taxa of the family *Flavobacteriaceae* and emended description of the family. *Int J Syst Evol Microbiol* **52**: 1049–1070.
- Bielawski, J., Dunn, K., Sabehi, G., and Beja, O. (2004) Darwinian adaptation of proteorhodopsin to different light intensities in the marine environment. *Proc Natl Acad Sci USA* **101**: 14824–14829.
- Brochier, C., Forterre, P., and Gribaldo, S. (2004) Archaeal phylogeny based on proteins of the transcription and translation machineries: tackling the *Methanopyrus kandleri* paradox. *Genome Biol* **5**: R17.
- Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.-A., Barrell, B.G., and Parkhill, J. (2005) ACT: the Artemis comparison tool. *Bioinformatics* **21**: 3422–3423.
- Cashmore, A.R., Jarillo, J.A., Wu, Y.-J., and Liu, D. (1999) Cryptochromes: blue light receptors for plants and animals. *Science* **284**: 760–765.
- DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.U., *et al.* (2006) Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496–503.
- Essen, L.O. (2006) Photolyases and cryptochromes: common mechanisms of DNA repair and light-driven signaling? *Curr Opin Struct Biol* **16**: 51–59.
- Frigaard, N.U., Martinez, A., Mincer, T.J., and DeLong, E.F. (2006) Proteorhodopsin lateral gene transfer between marine planktonic Bacteria and Archaea. *Nature* **439**: 847–850.
- Galagan, J.E., Calvo, S.E., Borkovich, K.A., Selker, E.U., Read, N.D., Jaffe, D., *et al.* (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422**: 859–868.
- Giovannoni, S.J., Tripp, H.J., Givan, S., Podar, M., Vergin, K.L., Baptista, D., *et al.* (2005a) Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**: 1242–1245.
- Giovannoni, S.J., Bibbs, L., Cho, J.-C., Stapels, M.D., Desiderio, R., Vergin, K.L., *et al.* (2005b) Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438**: 82–85.
- Giuliano, G., Al-Babili, S., and von Lintig, J. (2003) Carotenoid oxygenases: cleave it or leave it. *Trends Plant Sci* **8**: 145–149.
- Glockner, F.O., Kube, M., Bauer, M., Teeling, H., Lombardot, T., Ludwig, W., *et al.* (2003) Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. *Proc Natl Acad Sci USA* **100**: 8298–8303.
- Hallam, S.J., Putnam, N., Preston, C.M., Detter, J.C., Rokhsar, D., Richardson, P.M., and DeLong, E.F. (2004) Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* **305**: 1457–1462.
- Hallam, S.J., Mincer, T.J., Schleper, C., Preston, C.M., Roberts, K., Richardson, P.M., and DeLong, E.F. (2006) Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine *Crenarchaeota*. *PLoS Biol* **4**: 520–536.
- Hitomi, K., Okamoto, K., Daiyasu, H., Miyashita, H., Iwai, S., Toh, H., *et al.* (2000) Bacterial cryptochrome and photolyase: characterization of two photolyase-like genes of *Synechocystis* sp. PCC6803. *Nucleic Acids Res* **28**: 2353–2362.
- Jung, K.-H., Trivedi, V.D., and Spudich, J.L. (2003) Demonstration of a sensory rhodopsin in eubacteria. *Mol Microbiol* **47**: 1513–1522.
- Kajiwara, S., Fraser, P.D., Kondo, K., and Misawa, N. (1997) Expression of an exogenous isopentenyl diphosphate isomerase gene enhances isoprenoid biosynthesis in *Escherichia coli*. *Biochem J* **324**: 421–426.
- Kaneda, K., Kuzuyama, T., Takagi, M., Hayakawa, Y., and Seto, H. (2001) An unusual isopentenyl diphosphate isomerase found in the mevalonate pathway gene cluster from *Streptomyces* sp. strain CL190. *Proc Natl Acad Sci USA* **98**: 932–937.
- Kaneko, T., Nakamura, Y., Wolk, C.P., Kuritz, T., Sasamoto, S., Watanabe, A., *et al.* (2001) Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res* **8**: 205–213.
- Koonin, E.V., Makarova, K.S., and Aravind, L. (2001) Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* **55**: 709–742.
- Kuzuyama, T., and Seto, H. (2003) Diversity of the biosynthesis of the isoprene units. *Nat Prod Rep* **20**: 171–183.
- Man, D., Wang, W., Sabehi, G., Aravind, L., Post, A.F., Massana, R., *et al.* (2003) Diversification and spectral tuning in marine proteorhodopsins. *EMBO J* **22**: 1725–1731.

- Mongodin, E.F., Nelson, K.E., Daugherty, S., DeBoy, R.T., Wister, J., Khouri, H., *et al.* (2005) The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. *Proc Natl Acad Sci USA* **102**: 18147–18152.
- Nakamura, Y., Kaneko, T., Sato, S., Mimuro, M., Miyashita, H., Tsuchiya, T., *et al.* (2003) Complete genome structure of *Gloeobacter violaceus* PCC 7421, a cyanobacterium that lacks thylakoids. *DNA Res* **10**: 137–145.
- Nesbo, C.L., Boucher, Y., Dlugosz, M., and Doolittle, W.F. (2005) Lateral gene transfer and phylogenetic assignment of environmental fosmid clones. *Environ Microbiol* **7**: 2011–2026.
- Okamoto, O.K., and Hastings, J.W. (2003) Novel dinoflagellate clock-related genes identified through microarray analysis. *J Phycol* **39**: 519–526.
- Omelchenko, M.V., Makarova, K.S., Wolf, Y.I., Rogozin, I.B., and Koonin, E.V. (2003) Evolution of mosaic operons by horizontal gene transfer and gene displacement *in situ*. *Genome Biol* **4**: R55.
- Peck, R.F., Echavarri-Erasun, C., Johnson, E.A., Ng, W.V., Kennedy, S.P., Hood, L., *et al.* (2001) *brp* and *blh* are required for synthesis of the retinal cofactor of bacteriorhodopsin in *Halobacterium salinarum*. *J Biol Chem* **276**: 5739–5744.
- Ruiz-Gonzalez, M.X., and Marin, I. (2004) New insights into the evolutionary history of type 1 rhodopsins. *J Mol Evol* **58**: 348–358.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944–945.
- Sabehi, G., Massana, R., Bielawski, J.P., Rosenberg, M., DeLong, E.F., and Beja, O. (2003) Novel proteorhodopsin variants from the Mediterranean and Red Seas. *Environ Microbiol* **5**: 842–849.
- Sabehi, G., Beja, O., Suzuki, M.T., Preston, C.M., and DeLong, E.F. (2004) Different SAR86 subgroups harbour divergent proteorhodopsins. *Environ Microbiol* **6**: 903–910.
- Sabehi, G., Loy, A., Jung, K.-H., Partha, R., Spudich, J.L., Isaacson, T., *et al.* (2005) New insights into metabolic properties of marine bacteria encoding proteorhodopsins. *PLoS Biol* **3**: e273.
- Sandmann, G. (1994) Carotenoid biosynthesis in microorganisms and plants. *Eur J Biochem* **223**: 7–24.
- Sandmann, G. (2002) Molecular evolution of carotenoid biosynthesis from bacteria to plants. *Physiol Planta* **116**: 431–440.
- Sedkova, N., Tao, L., Rouviere, P.E., and Cheng, Q. (2005) Diversity of carotenoid synthesis gene clusters from environmental *Enterobacteriaceae* strains. *Appl Environ Microbiol* **71**: 8141–8146.
- Sineshchekov, O.A., Jung, K.-H., and Spudich, J.L. (2002) Two rhodopsins mediate phototaxis to low- and high-intensity light in *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci USA* **99**: 8689–8694.
- Spudich, J.L., Yang, C.S., Jung, K.H., and Spudich, E.N. (2000) Retinylidene proteins: structures and functions from archaea to humans. *Annu Rev Cell Dev Biol* **16**: 365–392.
- Suzuki, M.T., Beja, O., Taylor, L.T., and DeLong, E.F. (2001) Phylogenetic analysis of ribosomal RNA operons from uncultivated coastal marine bacterioplankton. *Environ Microbiol* **3**: 323–331.
- Suzuki, M.T., Preston, C.M., Beja, O., de la Torre, J.R., Steward, G.F., and DeLong, E.F. (2004) Phylogenetic screening of ribosomal RNA gene-containing clones in bacterial artificial chromosome (BAC) libraries from different depths in Monterey Bay. *Microb Ecol* **48**: 473–488.
- Teeling, H., and Gloeckner, F.O. (2006) RibAlign: a software tool and database for eubacterial phylogeny based on concatenated ribosomal protein subunits. *BMC Bioinformatics* **7**: 66.
- de la Torre, J.R., Christianson, L.M., Beja, O., Suzuki, M.T., Karl, D.M., Heidelberg, J., and DeLong, E.F. (2003) Proteorhodopsin genes are distributed among divergent marine bacterial taxa. *Proc Natl Acad Sci USA* **100**: 12830–12835.
- Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.
- Wolf, Y.I., Rogozin, I.B., Kondrashov, A.S., and Koonin, E.V. (2001) Genome alignment, evolution of prokaryotic genome organization, and prediction of gene function using genomic context. *Genome Res* **11**: 356–372.

Supplementary material

The following supplementary material is available for this article online:

Fig. S1. Neighbour-joining tree of concatenated ribosomal proteins L1, L10, L11 and L7/L12 (based on a comparison of 479 parsimony informative characters from inferred amino acid sequences) indicating a putative alphaproteobacterial origin for fosmid HF10_12C08. Bootstrap support values from 1000 pseudoreplicates given for major nodes (neighbour-joining/maximum parsimony).

Fig. S2. Neighbour-joining tree of concatenated ribosomal proteins S6, S18 and L9 (based on a comparison of 388 parsimony informative characters from inferred amino acid sequences) indicating a putative betaproteobacterial origin for BAC EB0_41B09. Bootstrap support values from 1000 pseudoreplicates given for major nodes (neighbour-joining/maximum parsimony; --, unresolved node).

Fig. S3. Neighbour-joining tree of xylose isomerase (based on a comparison of 374 parsimony informative characters from inferred amino acid sequences) supporting planctomycetal origin for fosmid HF10_49E08. Bootstrap support values from 1000 pseudoreplicates given for major nodes (neighbour-joining/maximum parsimony; --, unresolved node).

Fig. S4. Neighbour-joining tree of RNA polymerase A' subunit (based on a comparison of 938 parsimony informative characters from inferred amino acid sequences) supporting putative euryarchaeal origin for fosmid HF10_25C11. Bootstrap support values from 1000 pseudoreplicates given for major nodes (neighbour-joining/maximum parsimony).

Fig. S5. Phylogenetic tree of *blh* genes (based on a comparison of 317 parsimony informative characters from inferred amino acid sequences), which catalyses the oxidative cleavage of β -carotene to form retinal biosynthesis, from a variety of rhodopsin-containing organisms. Green branches indicate halorhodopsin-containing organisms and red indicate marine proteorhodopsin-containing. Green boxes indi-

cate sequences of archaeal origin. Sequences obtained in this study marked in bold. Bootstrap support values from 1000 pseudoreplicates given for two major nodes (neighbour-joining/maximum parsimony).

This material is available as part of the online article from <http://www.blackwell-synergy.com>